**A Text Mining Analysis of Back-to-School Discussion During COVID-19**

Junhe Yang

Department of Learning Technologies

University of North Texas, USA

[junheyang@my.unt.edu](mailto:junheyang@my.unt.edu)

Juncheng Ding

Department of Computer Science and Engineering

University of North Texas, USA

junchengding@my.unt.edu

**Abstract:** Preparing for back-to-school is essential, and it is also complicated because it involves different groups of people, and they have different needs. Pandemics, like COVID-19, make the situation even more difficult. Better preparation during a pandemic needs a comprehensive understanding of the situation. However, few studies have provided such information, especially during a pandemic. Furthermore, traditional research methods can hardly address the problems. This paper used text mining approaches to interpret open Twitter discussions on back-to-school preparation during COVID-19 and generated an inclusive insight into the preparation. Specifically, we collected 26219 related tweets and analyzed the user profile descriptions to determine who was involved in the preparation. We also examined the most frequently retweeted tweets and conducted a topic modeling analysis to find people’s concerns. Our research provides a holistic view of preparation for back-to-school during the pandemic, and it could help better prepare for back-to-school in the future pandemic.

# Introduction

Preparing for back-to-school is essential for a good learning environment and high teaching quality. It is not an easy task because many different people are involved in the process and have different needs. The pandemics, such as COVID-19, make the preparation even more complicated. For example, people may worry about their safety, which is rarely a problem before.

To better prepare for back-to-school during a pandemic, we should have a holistic understanding of the complicated situation. Specifically, we need to know who is involved in the preparation and what people worry about. Previous research has studied the back-to-school preparation before regular school seasons in different aspects. The researchers revealed that school administrators (Tierney & Jun, 2001), educators (Cooper & Alvarado, 2006), and parents (Norvillitis & MacLean, 2010) are all involved in the preparation, but they have different concerns. Researchers also demonstrated that the concerns are different during the pandemic compared to those in regular time. Flack et al. (2020) showed that educators mostly focus on student well-being and access to the technologies and the Internet under the crises. Levinson et al. (2020) collected parents' concerns, demonstrating that parents worry about students' safety and health. In Schlesselman, Cain, and DiVall's study (2020), students were concerned about the possibility of being infected and unafforded tuitions. However, these studies focused on only an individual group of people involved in the preparation, and few studies provide us an overall view of the preparation during the pandemic that includes all parties and their concerns.

Traditional research methods, such as the survey or the interview, can hardly generate a holistic view of the back-to-school preparation during the pandemic for two reasons. Firstly, they must target groups of people (e.g., students or teachers) before designing the study, and they cannot explore which groups are involved if not knowing ahead. Secondly, traditional study methods can hardly cover too many people due to the high cost, introducing significant bias. For example, if we study ten people and all of them are students or educators, we may conclude that only students and educators are concerned about preparing back-to-school, which may not be the fact. It is desirable to employ open discussions on the topic to generate a concise overview of back-to-school preparation during a pandemic.

To fill the above gaps, we used Twitter, which is one of the most active social media and includes open and informative discussions on the COVID-19 pandemic (Chen, Lerman, and Ferrara, 2020), to study who is involved in the back-to-school preparation and what their concerns are during the pandemic. We collected thousands of tweets on back-to-school and used different text mining techniques for the overview. Specifically, we analyzed the user profile descriptions to study who is involved in the preparation. Furthermore, we examined the most frequently retweeted tweets for the most concerned topics and conducted a topic modeling analysis for all the aspects in the preparation.

**Literature Review**

**Back-to-School Preparation During a Pandemic**

Recent research has studied different aspects of preparing for safety and healthy back-to-school amid the COVID-19 pandemic (Viner et al., 2020; Edmunds, 2020). For example, Daniel (2020) made guidelines in preparing for the new semester amid the pandemic. The research highlights the learning strategy as one of the concerns during the preparation in educators and school administrators' perspectives. Rega & Fink (2014) trained students on safe and healthy back-to-school in a pandemic simulation and collected their viewpoints. The research, which focused on students' perspective and the safety issue, demonstrated that students care about self-protect and health care before the new semester starts. Besides, Downes et al. (2020) and Panovska-Griffiths et al. (2020), following the line of the student's perspective and the safety issue in preparing for back-to-school, have studied the strategies such as testing and isolation of infected people during the COVID-19. These studies provided us information to prepare for back-to-school during a pandemic.

Nevertheless, these studies only focused on single issues and unique groups of people for their concerns, which is not enough for good back-to-school preparation. Khalifa et al. (2020) argued that a comprehensive overview is significant for preparing for back-to-school to diminish the risks under the crisis. Similarly, Darling-Hammond et al. (2020) claimed that people need a new and integrative view to support back-to-school in the long term. Consequently, it is necessary to provide a complete overview of who is involved and the issues in back-to-school preparation during a pandemic.

**Twitter Data and Its Text Mining Based Analytics**

Twitter is a microblogging platform involving different roles of users discussing various issues (Pak and Paroubek, 2010). Twitter contains information such as user profiles, hashtags, tweets, retweets, and so on, both of which are in the form of raw text. Previous research has successfully extracted various useful information from Twitter discussions (Java et al., 2007). For example, researchers employed Twitter in finding the public's concerns to provide authentic and effective information to health professionals during a health crisis from Twitter chat (Glowacki et al., 2016; Lazard et al, 2015). Following such research, we applied text analysis on Twitter’s back-to-school discussions aiming to generate a holistic view of back-to-school preparation.

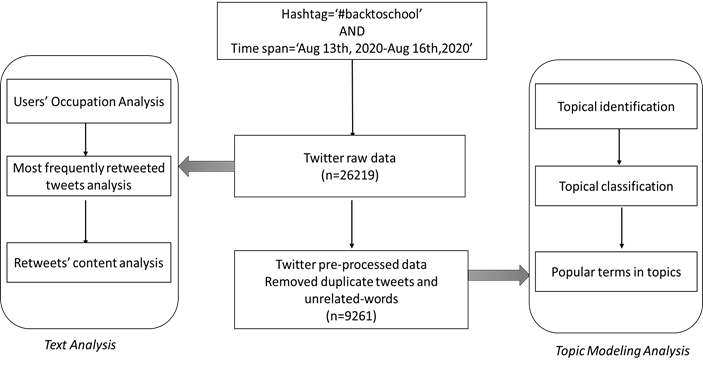
Researcher have developed different text mining-based approaches to analyze Twitter texts. For example, Vainio and Holmberg (2017) used Twitter users’ profile descriptions to analyze the users’ demographic information. Besides, recent study analyzed most retweeted tweets to find what topics attracted the most attention during the pandemic (Kaminski et al., 2020). Moreover, Doogan et al., (2020) conducted topic modeling analysis to investigative the public’s perspectives and attitudes regarding COVID-19 pandemic. In this paper, we conducted such analyses to find who is discussing back-to-school and what people’s concerns are. Specifically, this study emphasized on the following research questions:

(1) Who involves in the back-to-school preparation amid the COVID-19 pandemic?

(2) What are the major issues during the preparation?

**Methodology**

Following the previous research, this study collected tweets that contained the hashtag #backtoschool in a time span before school starts. After collecting the tweets, we categorized the people involved in the discussion into different groups and analyzed the most prevailing tweets in such discussion using text mining-based approaches. Finally, this study conducted the topic modeling analysis to examine the major topics of back-to-school. We present the flowchart to illustrate the data collection and analysis methodologies in Figure. 1. Generally, the analyses include three parts: data retrieval and preprocessing, users and trending tweets analysis, and topic modeling analysis.



**Figure 1.** The Flowchart of Data Collection and Analysis Methodology

**Data Collection and Preprocessing**

Following the prior study (Highfield, Harrington & Bruns, 2013), we described our topic back-to-school as the keyword “backtoschool” and used the respective specific hashtag “#backtoschool” to gather related tweets before the new semester started. Specifically, we collected the tweets that include “#backtoschool” between 13th and 16th August 2020 using python and Twitter Application Programming Interface (API), and we chose the time because it was the date before many schools start in the US. The assumption is that the collected tweets were discussing back-to-school by different people involved in the process. We only kept the English tweets in our raw data and finally got 26219 raw tweets from 13278 different Twitter users containing in a JSON document. Each piece of tweet data contains its main text, whether it is a retweet, the user’s profile description, et al. We used the user profile description and the retweeting information to categorize the users and identify the trending issues, respectively. We conducted the topic modeling analysis using the tweets’ main text after preprocessing.

We removed duplicate tweets and meaningless words inside tweets in preprocessing the text for our topic modeling analysis. A tweet may be retweeted, resulting in duplicate tweets with the same content in our dataset. In our preprocessing, we just kept the original content of all the duplicates and removed the retweets. Consequently, we got 9261 tweets, nearly 35% of the total tweets (n=26219) for the topic modeling analysis. Besides, we used Regular Expressions to split the text into unique words and removed all “stopwords” (e.g., “I”, “me”, “they”, “he”, “she”, et al.) that are meaningless to content analysis numbers, and symbols, tags, URLs, hashtags before topic modeling following the standard approach in topic modeling analysis. We also set all the uppercase letters to lowercase ones.

**Users and Trending Tweets Analysis**

In this part of analysis, we categorized the users involved in this discussion to identify who is involved in the back-to-school preparation and analyzed the most trending tweets for the prevailing issues in back-to-school preparation. The assumption behind is that most people who actively join in the discussion are involved in the back-to-school preparation and that the most trending tweets (having the greatest number of retweets) in the discussion reveal the most prevailing issues.

We conducted text mining-based approaches to decode the users' profile descriptions. We gathered 13278 users’ profile descriptions to identify their occupations in this analysis. Specifically, we specified several keywords to cluster similar roles in one category, i.e., if a user whose profile description contains a specific keyword, we consider the user as the respective occupation. For example, the keywords "professor", "teacher", "educator", "instructor", and "coach" are for educators. We also created an “other” category for the users with non-educational roles, such as the photographer, designer, blogger, and users without profile descriptions. By doing so, we concluded the users into seven occupational categories regarding their roles in the back-to-school preparation: educators, organizations, parents, students, administrators, education-related workers, and others.

We conducted a statistical analysis on the retweeted tweets and examined the most retweeted ones to find the more prevailing issues. This study only listed the original tweets among their many retweets.

**Topic Modeling Analysis**

It is desirable to summarize topics that exist in the tweets to generate a high-level view of the discussion, but it is extremely hard to read the thousands of tweets and conclude the topics manually. The topic model provides a solution to detect latent topics existing in the text. Blei et al. (2013) developed the most recognized topic model, named Latent Dirichlet Allocation (LDA), to generate a topic-level view of a large text collection. We used LDA to generate a holistic view, described by several topics, of the content within our collected tweet data. Furthermore, we analyzed the topics to find the major concerns during the COVID-19 back-to-school discussion.

**Results**

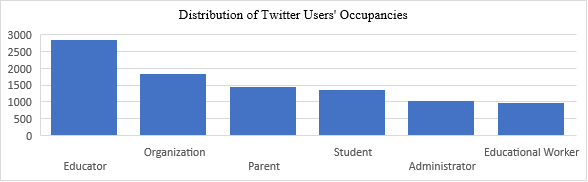
**Twitter Users’ Profile Descriptions Analysis**

Researchers have shown that we can use the user profile descriptions to summarize the user’s demographic information (Vainio and Holmberg, 2017). Similarly, we can also safely use such descriptions to categorize the users by their occupations. This study analyzed all 13278 user’s profile descriptions. As explained before, we clustered the users involved in the discussion into seven categories according to several keywords. Table 1 shows the details in our clustering that include the occupation categories, keywords, profile descriptions, and examples.

We argue that many people in a group will discuss a topic if the group cares more about it. Therefore, we drew the number of people involves in each of the seven occupational categories to see their different levels of concern. Specifically, we generated a distribution of the number of people in each occupation extracted from the profile descriptions (See Figure 2). In our sampled Twitter users, “Educator” (n=2860) has the largest portion. 1850 users are for “Organization” as the second most popular group. The numbers of “Parent” and “Student” are 1457 and 1377, as the third and fourth largest groups. Following the two groups is the “Administrator” which accounts for 1030 users. The last group is “Educational Worker” with the number of 991.

| Category | Keywords | Examples of Descriptions in user profile |
| --- | --- | --- |
| Educator | “professor”, “teacher”, “educator”, “lecturer”, “instructor”, “coach” | “Assoc. Prof @usfmuma researching #culture x #gender x #entrepreneurship. #millennial 🇨🇺🇺🇸 Made in Miami.”  “Husband, father, educator. Be humble and kind......... please.” |
| Organization | “org”, “school”, “institute”, “univ”, “college”, “community” | “The community for schools in Europe: eTwinning is co-funded by the Erasmus+ programme of the European Union and is managed in the UK by the @BritishCouncil.”  “Kuwait University”  “Healthy Schools Network, Inc. is a 501 c3 national environmental health organization. Founded in 1995.” |
| Parent | “parent”, “family”, “mom”, “mother”, “dad”, “father” | “#BlackLivesMatter  Sing. Love. Dance. | Husband. Father. Family. Techie 💻, Gamer 🕹️, Geek 🤓. Raptors, Blue Jays, RomComs, Web, Outdoors & #thelittlethings.”  “Husband, Father, Son, Brother, Uncle, Sigma Chi. Works at @wellfleetins but tweets are my own. #GoMocs #LetsGoX #YNWA #HereWeGo #LakeShow #GoCFC” |
| Student | “learner”, “student”, “kid”, “children”, “phd” | “believer, wife, learner, bibliophile, @ThinkingMapsTrainer, #DigCit, Instructional Technology Specialist-making the most of every day!”  “Wife, book nut, F1 fan, forever student! Limerick in TO! Still do not know what I want to be when I grow up! Never met an egg I did not like!” |
| Administrator | “coordinator”, “director”, “administrator”, “founder”, “leader”, “president” | “Speaker, Entrepreneur, Mkt & Media Expert, Christian Minister & Artist. I love God, my wife, my daughters & try my best to honor them. Jesus is my inspiration!”  “Mental Health advocate & speaker and host of Cheryl Panics podcast. Founder of @AnxietyAide and @PanicAide natural 2oz shots. Lover of Christ. Wife. Mom to 4.” |
| Educational Worker | “advisor”, “specialist”, “librarian”, “consultant”, “technicism” | “Librarian on wheels! CoFounder of Books on Bikes and school librarian”  “STEM Specialist. Science Enthusiast. ΔΣΘ  https://t.co/Z7g07wkMuq” |
| Others | unrelated with educational area, like photographer, designer, sociologist, blogger, and so on. Unverified users, like adding images, emoticon, or motto in the description. | “🤔✌🏻”  “🇨🇦 🌈✊🏿😷Maker. Building a sustainable home+B&B in #princeedwardcounty. Otherwise making software, droning, kayaking. Work: @marketcircle, x-@flixel”  “Toronto's source for local news and culture, restaurant reviews, event listings and the best of the city. Subscribe to our daily podcast Only in Toronto.” |

**Table 1.** Occupation Categories, Keywords, Descriptions, and Examples



**Figure 2.** Users’ Occupancies of Distributions in Categories

**Most Frequently Retweeted Tweets Analysis**

We argue that the more a tweet is retweeted, the more people care about the tweet’s topic. In this regard, we found ten most frequently retweeted tweets to identify the most concerning topics in back-to-school discussion. Table 2 lists such tweets with its number of retweets, its content, and its user information. We examined these retweeted tweets to find the most prevailing topics. This analysis informs us of the most concerned topics during back-to-school in COVID-19.

| Retweeted Tweets No. | Retweeted Count | Most Frequently Retweeted Tweet’s content | User’s Type | User’s Name |
| --- | --- | --- | --- | --- |
| No.1 | 475 | “FOLLOW &amp; TAG a mate with #WinStudio in the comments for a chance to WIN a pair of Kickers.” | Organization | @studio\_co\_uk |
| No.2 | 449 | “If your classroom is virtual, blended, or in-person this school year, try creating a digital #BackToSchool bulletin board” | Organization | @GoogleForEdu |
| No.3 | 318 | “School ventilation🌬️could spread #COVID19. Why aren’t we talking about it? by @alexbozikovic https://t.co/yYN4ZN9fcd v… #backtoschool" | Non-related educational people | @picardonhealth |
| No.4 | 308 | “Today, I am proud to launch the #BackToSchool Campaign.  The Govt has lied and cheated its way through the last 5 months. Millions of children have been deprived of an education since March.  Join me and show the Govt we will not be accepting this new normal for our children.” | Non-related educational people | @simondolan |
| No.5 | 278 | “#BackToSchool #GSuiteEdu tip: your students can easily follow along on #GoogleSlides with real-time closed captions power” | Organization | @GoogleForEdu |
| No.6 | 263 | “It is requested to competent authorities to facilitate our return back to China. To continue our research because we have wasted half year.  #TakeUsBackToSchool” | Student | @AyeshaK68608341 |
| No.7 | 183 | “Return to school a 'false choice' for many families. For low-income families and essential workers, online learning is out of the question. via  #Covid19 #sdoh #BackToSchoolSchool” | Non-related educational people | @Andr\u00e9 Picard |
| No.8 | 181 | “The focus should be when and how, not whether school should resume” | Educator | @Krista B-W |
| No.9 | 162 | “Ontario Catholic school board says teachers take unpaid leave, vacation time, resign or retire if they refuse” | Non-related educational people | @signsaresaying |
| No.10 | 133 | “We want your child to start the 2020-2021 academic year successfully! @MDCPS has developed a #BackToSchool checklist” | Organization | @Miami-Dade Schools |

**Table 2.** Top 10 Frequently Retweeted Tweets’ Analysis

**Topic Modeling Analysis**

To find all the topics in the discussion, we used preprocessed data (explained in the data preparation section) and the LDA model for conducting the topic modeling analysis. We set the number of topics as 8, 10, 20, 40 to find the best parameter. We found that LDA with 20 topics can best interpret the back-to-school discussion and presented the 20 topics in Table 3. To better analyze the topics, we grouped the 20 topics into four categories: sharing information and opinions, teachers’ and students’ health, teaching strategies and technologies, and supplies shopping.

|  |
| --- |
| *Sharing information and opinions*  Topic 0: first, day, new, episode, students, made, support, parents, colleges, information, addition, collection  Topic 1: blog, teaching, super, choose, post, reduce, awesome, learning, interesting, edtech. Fantastic, tips  Topic 2: free, webinar, would, teachers, thread, strategies, education, planning, healthy, work, resources, series  Topic 3: wait, can’t, open, parents, reopening, schools, sign, teacher, safe, working, meet, students  Topic 4: districts, impact, changes, sharing, questions, form, covid, asking, guide, delivery, return, plans, clinic  Topic 5: edchat, reopening, video, help, stay, making, solution, send, schoolopening, notebook, need, schools*.* |
| *Health*  Topic 6: air, give, away, change, strategies, space, help, guide, soft, open, numbers, classroom  Topic 7: reopen, way, blog, post, new, wearing, academic, better, kids, colleagues, get, masks, looking  Topic 8: work, focus, edtech, colleges, cleaning, virus, support, schools, help, students, public, entered  Topic 9: mask, face, supporting, daily, facemask, child, cotton, handmade, keeping, layer, fitted, size |
| *Teaching strategies and technologies*  Topic 10: covid, there’s, everything, virtually, always, cool, important, shared, giveaway, need, guidance, children  Topic 11: remotelearning, get, let, students, physical, importance, resources, model, practices, visit, per, help  Topic 12: school, away, homeschool, new, style, question, packs, start, choose, giving, teacher, models  Topic 13: point, hybrid, spread, prepared, perfect, ideas, kids, build, issues, issue, wear, masks  Topic 14: coronavirus, right, going, classes, person, follow, support, learning, next, buildings, new, session  Topic 15: new, ready, students, welcome, staff, looking, getting, forward, googleclassroom, already, unique, help  Topic 16: classroom, home, prepare, social, started, new, kids, learn, great, teacher, media |
| *Supplies Shopping*  Topic 17: friends, know, essentials, group, orders, looking, adidas, child, getting, new, need, lovely  Topic 18: great, anything, coming, small, system, opportunity, free, shows, new, join, supplies, beautiful  Topic 19: mens, cute, space, study, styles, shopping, ralph, lauren, classic, fantastic, polo, concerned  Topic 20: nike, boys, business, must, fitness, girls, long, access, short, plans, workout, running |

**Table 3.** All 20 Topics in Back-to-School Discussion

**Discussion**

We collected tweets discussing back-to-school before the new semester started in COVID-19 to generate a review of the back-to-school preparation during the pandemic. The findings showed that different groups of people were involved in the back-to-school preparation and they had different perspectives and worries regarding the preparation.

**Different Groups of People in Back-to-School Preparation**

We analyzed who are involved in the preparation via their involvement in the open discussion. We can observe from Figure 2 that educators accounted for the largest portion among all the occupancies, which indicates that they were mostly worried about preparation for back-to-school during the pandemic. One reason may be that educators would like to use education-related Twitter hashtags to share insights and discuss related topics, consistent with the findings of Rosenberg et al.’s (2016) and Carpenter et al.’s studies (2020). Another two large groups were parents and students. They paid great attention to planning back-to-school in the continuing COVID-19 pandemic as in typical times. Research has already investigated these three groups during back-to-school preparation, either during pandemics or regular times. Our research has discovered that groups other than the above three are also involved in the preparation.

Different from previous studies, this study found that many organizations were actively involved in preparation for back-to-school. To control the spread of the pandemic, organizations such as the World Health Organization (WHO), United Nations Educational, Scientific, and Cultural Organization (UNESCO) provided professional advice to help schools reopen. Although few prior studies demonstrated organizations helped prepare for back-to-school, this finding showed that organizations are actively involving in the back-to-school preparation by spreading useful information and professional guidelines. Our study also indicates that educational administrators and employees are also actively involved in preparing for returning the school during the pandemic, which is rarely discussed before. They could be that they cared about the health and safety issues that were not that significant before. Besides the individuals and organizations closely related to education, those from other fields, like photographers, designers, and sociologists, were also involved in preparing this issue during the pandemic. These results cast new insights into prior literature: other than educators and scholars who worry about returning the school in teaching and learning aspects (Landeros et al., 2020; Stephenson, 2020), other groups (i.e., health and educational organizations, educational workers, and others not related to the education) are also actively involving into the preparation for back-to-school during the pandemic. To conclude, a better back-to-school preparation during pandemics needs to consider all these groups.

**Most Retweeted Tweets in Back-to-School Discussion**

Knowing who are involved in the preparation is not enough. We also found the most prevailing topics via the most retweeted tweets, as in Table 2. The most retweeted tweet (No. 1) was about getting free essentials from online shopping websites before the new semester started, which is the same as in regular back-to-school seasons. Many merchants sent gifts to the people, as usual, to attract them to buy supplies before the new semester started. Even though buying supplies for back-to-school is still the main preparation for the people under the pandemic situation.

Besides finding the school supplies topic, this study also revealed that the topic (No. 3) that the school ventilation could result in the spread of COVID-19 attracted great public discussion. The reason could be that the ventilation issue contrasted with scientific reports showing that school ventilation could stop the spread of COVID-19 (Gray, 2020; Fauci, 2020). This topic reminded people to pay attention to whether school ventilation can prevent the spread of COVID-19, which is a severe problem of the back-to-school preparation in the pandemic.

Another critical topic (No. 2, No. 4, and No. 5) is that people were still concerned about using innovative technologies and new class formats amid the COVID-19 pandemic, consistent with the previous studies that demonstrated it was time to use innovative digital tools to deliver the knowledge, organize the academic activities, and increase the workplaces' flexibility during the pandemic (Schwarz et al., 2020). These findings suggest that people who share or discuss back-to-school tended to prepare a more safe and effective learning environment than usual for students and educators. However, this study also reveals some new aspects in the preparation.

Many people had expressed desire and need for back-to-school in the coming semester (No. 6 and No. 7). Students desired to go back-to-school while students from low-income families needed to go back to school due to financial reasons. The reason could be that letting low-income families' children back to school can reduce the parents' pressures and help them put effort into the new "normal" life. These findings showed that people in different roles desired to reopen the school and return to the "new" normal. To keep students and academic staff safe and healthy back to school, the public focused on the government's tweets to guide which schools and districts can be safely reopened (No. 8, No. 9, and No. 10). The public's concern was consistent with scholars about when, how, or whether the school can reopen under a specific situation. Therefore, this part of the study suggests that although some people had a strong desire to back-to-school amid the COVID-19 pandemic, they need to base professional information from organizations to have adequate preparation before school reopened.

**Hot Topics in Back-to-School Discussion**

Knowing all the topics other than only the prevailing ones is essential for a complete understanding of the preparation. We analyzed all the 20 topics in the discussion extracted by LDA in this part. We summarized all the 20 topics into four aspects of back-to-school preparation in Table 3.

One category of topics is sharing information and resources. The topics show that people were concerned about collecting and sharing useful information and professional support for back-to-school preparation. Topics in this category discuss different aspects of the issue. For example, this category's topic showed that online webinars, chats, or blogs were popular ways to help post educational information or resources to the public. Another topic revealed that it is necessary to share information and resources because different schools had different needs. Several (No. 4, No. 8, and No.10) of the most retweeted tweets also falls into this category of topics. Therefore, sharing useful information based on urgent and specific needs and providing ongoing support is essential for back-to-school preparation during a pandemic.

The second category of topics is about students' and educators' health. Unlike in regular times, people's health is a significant premise for back-to-school during the pandemic. This category includes topics on the ventilation (also found in the No. 3 most retweeted tweet), the strategy to keep sanitary, encouraging students' exercises, the mandate to wear masks, et al. Back-to-school preparation during a pandemic should pay great attention to the health issue.

Teaching strategy and technology is the third category of topics, which also accounts for a large part of the preparation. In this complicated situation, different schools have changed their class format and teaching strategies over time. Our findings also showed that remote learning is a good teaching model that can help students keep the social distance. Meanwhile, we also found that a new class format, named hybrid classrooms, was a popular teaching strategy during the pandemic. Besides, many families are more interested in a new learning style, homeschool, in which students can stay at home to continue learning. The most retweeted tweet No. 2 and No. 5 belong to this category. Overall, it is crucial to consider different teaching strategies for maintaining the quality of teaching and learning in an unprecedented time.

The last category of topics is shopping for back-to-school supplies. This category of topics includes the No. 1 most retweeted tweet. Like typical back to school season, the public always focuses on making shopping lists of buying essentials. Some people would like to share their essentials' orders or discuss it with their friends. Some companies provide free supplies to attract people to shopping for back-to-school preparation. Hence, future preparation in the pandemic period should also take action to guarantee that students have an abundance of supplies successfully returning to the school during the pandemic.

To conclude, our topic modeling analysis provides a complete overview of the topics in back-to-school discussion during the pandemic. We argue that this is a complete one because we generate such an overview from a large-scale and open discussion. Our overview could serve as guidelines for future back-to-school preparation during a pandemic.

**Limitations**

This study has some limitations. Firstly, we only chose English tweets. Tweets in other languages may provide different insights about back-to-school during the COVID-19 pandemic. Therefore, future studies can consider using tweets in different languages to analyze different countries or locations people's concerns and opinions regarding back-to-school.

Besides, we used LDA, a classic topic modeling tool, to identify potential tweets' topics, while other methods may offer different types of functions and analysis. Hence, future studies can use sentiment analysis to systematically analyze the public's attitudes about returning to the school amid the COVID-19.

**Conclusion**

Our study provides new insights into preparation back-to-school during the COVID-19 pandemic. This study shows that other than the educators, parents, and students in the current research, organization, educational worker, and people not related to education are also involved in the back-to-school preparation. We also revealed the ten most concerning issues during the preparation. People have different and various concerns about the back-to-school preparation. Therefore, we extracted 20 issues on their concerns and categorized the concerns into four aspects: sharing educational information and resources, students' and educators' health, different teaching strategies and technologies, and school essentials shopping. Our research generates a holistic overview of the back-to-school preparation during COVID-19. We believe our research will help prepare for future back-to-school preparation during a pandemic.

**References**

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communication, 13*(1), 210-230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. *JMIR Public Health and Surveillance, 6*(2), e19273. https://doi.org/10.2196/19273

Cooper, J. M., & Alvarado, A. (2006). *Preparation, recruitment, and retention of teachers*. Brussels: International Institute for Educational Planning.

Carpenter, J. P., Willet, K. B. S., Koehler, M. J., & Greenhalgh, S. P. (2020). Spam and educators’ Twitter use: Methodological challenges and considerations. *TechTrends, 64*(3), 460-469. https://doi.org/10.1007/s11528-019-00466-3

Daniel, S. J. (2020). Education and the COVID-19 pandemic. *Prospects*, 1-6. https://doi.org/10.1007/s11125-020-09464-3

Darling-Hammond, L., Schachner, A., Edgerton, A. K., Badrinarayan, A., Cardichon, J., Cookson Jr, P. W., ... & Martinez, M. (2020). Restarting and Reinventing School: Learning in the Time of COVID and Beyond. *Journal of the Pediatric Infectious Diseases Society.* https://doi.org/10.1093/jpids/piaa095

Downes, K. J., Danziger-Isakov, L. A., Cousino, M. K., Green, M., Michaels, M. G., Muller, W. J., ... & Ardura, M. I. (2020). Return to School for Pediatric Solid Organ Transplant Recipients in the United States During the COVID-19 Pandemic: Expert Opinion on Key Considerations and Best Practices. *Journal of the Pediatric Infectious Diseases Society*. https://doi.org/10.1093/jpids/piaa095

Doogan, C., Buntine, W., Linger, H., & Brunt, S. (2020). Public Perceptions and Attitudes Toward COVID-19 Nonpharmaceutical Interventions Across Six Countries: A Topic Modeling Analysis of Twitter Data. *Journal of Medical Internet Research, 22*(9), e21419. https://doi.org/10.2196/21419

Edmunds, W. J. (2020). Finding a path to reopen schools during the COVID-19 pandemic. *The Lancet Child & Adolescent Health.* https://doi.org/10.1016/S2352-4642(20)30249-2

Fauci, A., (2020). COVID-19 study: School ventilation systems could play role in stopping spread, *Eyewitness News,* <https://abc7chicago.com/school-ventilation->covid-coronavirus-schools-covid-19-spread-study/6342910/

Gray, L., (2020). Ventilation expert reveals how Texas schools can stop the spread of COVID in classroom, *Houston Chronicle*, <https://www.houstonchronicle.com/life/article/Texas-school->ventilation-expert-covid-class-15542882.php

Glowacki, E. M., Lazard, A. J., Wilcox, G. B., Mackert, M., & Bernhardt, J. M. (2016). Identifying the public's concerns and the Centers for Disease Control and Prevention's reactions during a health crisis: An analysis of a Zika live Twitter chat. *American journal of infection control, 44*(12), 1709-1711. https://doi.org/10.1016/j.ajic.2016.05.025

Highfield, T., Harrington, S., & Bruns, A. (2013). Twitter as a technology for audiencing and fandom: The# Eurovision phenomenon. Information, *Communication & Society, 16*(3), 315-339. https://doi.org/10.1080/1369118X.2012.756053

Khalifa, S. A., Mohamed, B. S., Elashal, M. H., Du, M., Guo, Z., Zhao, C., ... & El-Seedi, H. R. (2020). Comprehensive Overview on Multiple Strategies Fighting COVID-19. *International Journal of Environmental Research and Public Health, 17*(16), 5813. https://doi.org/10.3390/ijerph17165813

Kamiński, M., Szymańska, C., & Nowak, J. K. (2020). Whose Tweets on COVID-19 Gain the Most Attention: Celebrities, Political, or Scientific Authorities?. *Cyberpsychology, Behavior, and Social Networking.* https://doi.org/ 10.1089/cyber.2020.0336

Levinson, M., Cevik, M., & Lipsitch, M. (2020). Reopening primary schools during the pandemic. *Medicine and Society.* https://doi.org/10.1056/NEJMms2024920

Lazard, A. J., Scheinfeld, E., Bernhardt, J. M., Wilcox, G. B., & Suran, M. (2015). Detecting themes of public concern: a text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. *American journal of infection control, 43*(10), 1109-1111. https://doi.org/10.1016/j.ajic.2015.05.025

Landeros, A., Ji, X., Lange, K. L., Stutz, T. C., Xu, J., Sehl, M. E., & Sinsheimer, J. S. (2020). An Examination of School Reopening Strategies during the SARS-CoV-2 Pandemic. *medRxiv*. https://doi.org/10.1101/2020.08.05.20169086

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, 56-65.

Norvilitis, J. M., & MacLean, M. G. (2010). The role of parents in college students’ financial behaviors and attitudes. *Journal of economic psychology, 31*(1), 55-63. https://doi.org/10.1016/j.joep.2009.10.003

Panovska-Griffiths, J., Kerr, C., Stuart, R. M., Mistry, D., Klein, D., Viner, R. M., & Bonell, C. (2020). Determining the optimal strategy for reopening schools, work and society in the UK: balancing earlier opening and the impact of test and trace strategies with the risk of occurrence of a secondary COVID-19 pandemic wave. *medRxiv*. https://doi.org/10.1101/2020.06.01.20100461

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *LREc* (Vol. 10, No. 2010, pp. 1320-1326).

Rega, P. P., & Fink, B. N. (2014). Immersive simulation education: a novel approach to pandemic preparedness and response. *Public Health Nursing, 31*(2), 167-174. https://doi.org/10.1111/phn.12064

Rosenberg, J. M., Greenhalgh, S. P., Koehler, M. J., Hamilton, E., &Akcaoglu, M. (2016). An investigation of state educational twitter Hashtags (SETHs) as affinity spaces. *E-Learning and Digital Media, 13*, 24–44. https://doi.org/10.1177/2042753016672351

Schlesselman, L. S., Cain, J., & DiVall, M. (2020). Improving and Restoring the Well-being and Resilience of Pharmacy Students during a Pandemic. *American Journal of Pharmaceutical Education, 84*(6).

Stephenson, J. (2020). National Academies Offers Guidance on Reopening Schools Amid COVID-19 Pandemic. *JAMA Health Forum* (Vol. 1, No. 7, pp. e200973-e200973). https://jamanetwork.com/channels/health-forum/fullarticle/2768951?utm\_source=silverchair&utm\_campaign=jama\_network&utm\_content=covid\_weekly\_highlights&utm\_medium=email

Schwarz, M., Scherrer, A., Hohmann, C., Heiberg, J., Brugger, A., & Nuñez-Jimenez, A. (2020). COVID-19 and the academy: It is time for going digital. *Energy research & social science, 68*, 101684. https://doi.org/10.1016/j.erss.2020.101684

Tierney, W. G., & Jun, A. (2001). A university helps prepare low income youths for college: Tracking school success. *The Journal of Higher Education, 72*(2), 205-225. https://doi.org/10.1080/00221546.2001.11778878

Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., ... & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *The Lancet Child & Adolescent Health*. https://doi.org/10.1016/S2352-4642(20)30095-X

Vainio, J., & Holmberg, K. (2017). Highly tweeted science articles: who tweets them? An analysis of Twitter user profile descriptions. *Scientometrics, 112*(1), 345-366. https://doi.org/10.1007/s11192-017-2368-0